First Day of CS203: **Advanced Computer Architecture**

Hung-Wei Tseng

CS203: Let's say something!

What's your name?

What's your favorite topic in computer science?



Why're you taking CS203

Google (How Often Do People Lie				
	Q All	E News	🖾 Images	▶ Videos	Shopping
	About 26	67,000,000 i	results (0.54 s	econds)	
The study, published in the jour at least once during a 10-minut	rnal's June issue, found t				
	lies. "Pe	eople tell a	considerabl	le number of	f lies in every
	UMas	s resear	cher finds	most peo	ple lie in e
	https://v	ww.eurek	alert.org > pu	ib_releases	> uoma-urf06
politics 45 Cor	ngress SCC	OTUS Facts	First 2020 20)19 Elections	
		L.			f t
Donald Trump lies more of				Itten	
hands	ev	ery	day		
sis by <mark>Chris Cillizza</mark> , CN ted 4:56 PM ET. Mon Ju	NN Editor-a ine 10. 2019	t-large 9			
	Google politics 45 Cor Ald Tru hands sis by Chris Cillizza, Cl ted 4:56 PM ET, Mon Ju	Google How O All About 20 The stu at least lies. "Pe UMass https://w Politics 45 Congress SCO Ald Trump hands ev sis by Chris Cillizza, CNN Editor-a ted 4:56 PM ET, Mon June 10, 2019	How Often Do Per All News About 267,000,000 r The study, publish at least once durin lies. "People tell a UMass researcy https://www.eureka Politics 45 congress SCOTUS Facts Add Trump lies bands every sis by Chris Cillizza, CNN Editor-at-large ted 4:56 PM ET, Mon June 10, 2019	How Often Do People Lie All News Images About 267,000,000 results (0.54 set The study, published in the jour at least once during a 10-minu lies. "People tell a considerable UMass researcher finds https://www.eurekalert.org > pu Politics 45 Congress SCOTUS Facts First 2020 20 Ald Trump lies mod bands every day sis by Chris Cillizza, CNN Editor-at-large ted 4:56 PM ET, Mon June 10, 2019	Google How Often Do People Lie All News Images Videos About 267,000,000 results (0.54 seconds) The study, published in the journal's June i at least once during a 10-minute conversa lies. "People tell a considerable number of UMass researcher finds most peo https://www.eurekalert.org > pub_releases politics 45 Congress SCOTUS Facts First 2020 2019 Elections ald Trump lies more of https://www.eurekalert.org > pub_releases About 267,000,000 results (0.54 seconds) Mass researcher finds most peo https://www.eurekalert.org > pub_releases ald Trump lies more of hands every day sis by Chris Cillizza, CNN Editor-at-large ted 4:56 PM ET, Mon June 10, 2019



CS203: Let's say something!

What's your name?

What's your favorite topic in computer science?



Why're you taking CS203

CS203: Let's say something!



What's your name?



The return of backpropagation

- Between 2005 and 2009 researchers (in Canada!) made several technical advances that enabled backpropagation to work better in feed-forward nets.
 - Unsupervised pre-training; random dropout of units; rectified linear units.
 - The technical details of these advances are very important to the researchers but they are not the main message.
 - The main message is that backpropagation now works amazingly well if you have two things:
 - a lot of labeled data
 - a lot of convenient compute power (e.g. GPUs)



2018 Turing Award

- Cha

Hung-Wei Tseng

David Patterson

John Hennessy



Sixth Edition

John L. Hennessy David A. Patterson

COMPUTER ARCHITECTURE

A Quantitative Approach



Computer Architecture



Enables

Deep Learning

Computer architecture also enables ...



What's computer architecture?

\bigcirc		JOIN MWU GAMES TRAVELER	BROWSE THESAURUS	WORD OF THE DAY
Merriam- Webster	n- SINCE 1828	architecture		
vvebsiel		DICTIONARY	THESAURUS	

architecture noun

ar-chi-tec-ture | Viär-kə-jtek-chər 🕥 🕅 Definition of archite The manner in which the components

- 1 : the art or science of building specifically: the art or of a computer or computer system are habitable ones
- ² a : formation or construction resulting from organized and integrated // the architecture of the garden
 - **b** : a unifying or coherent form or structure // a novel that lacks architecture
- 3 : architectural product or work *II* buildings that comprise the *architecture* of the square
- 4 : a method or style of building // Gothic architecture
- : the manner in which the components of a computer or computer system are 5 organized and integrated

// different program architectures



What're those "components"?

Desktop Computer



Server



 Peripher
 DRA

 als (e.g.,
 DRA

 GPUs)
 Horse

 Process
 Process

Peripherals (e.g.

I/O Connectors (e.g., keyboard/mouse)

als (e.g., DRAM DRAM DRAM DRAM

Processor Processor

DRAM DRAM DRAM DRAM

MacBook Pro 13"





iPhone 11





Play Station 4



Processor + GPU



Peripherals (e.g., H.D.D.)

Peripher als (e.g., codecs)

Nintendo Switch

(e.g., HDMI)

I/O Connectors

1 15

DRAM



Processor + GPU

<u>i</u>. •

Network Modules/ Codecs

Peripherals (e.g., memory cards.)

Tesla Model 3



Processors and memory modules are everywhere!







Why are "Processor" & "Memory" everywhere?

von Neumman Architecture



By loading different programs into memory, your computer can perform different functions







00005d24 00c2f000 nstructi ta 0000bd24 0000008 0 0 2ca422a0 00c2f800 80000008 130020e4 00003d24 00c30000 2ca4e2b3 80000008

Storage



Source Code

Compiler (e.g., gcc)

Program

0f00bb27 509cbd23 00005d24 0000bd24 2ca422a0 130020e4 00003d24 2ca4e2b3

Data

00c2e800 80000008 00c2f000 80000008 00c2f800 00000008 00c30000 80000008

How my "Java code" becomes a "program"



Compiler (e.g., javac)

Jave Bytecode (.class)

Source Code

Ê

lava

cafebabe 0000033 001d0a00 ta 06000f09 00100011 0800120a 00130014 07001507

00c2e800 00000008 00c2f000 80000008 00c2f800 0000008 00c30000 80000008

How my "Python code" becomes a "program"

Libraries



Source Code Python Perfector

Interpreter (e.g., python)

Program

0f00bb27 509cbd23 00005d24 0000bd24 2ca422a0 130020e4 00003d24 2ca4e2b3

Data

00c2e800 0000008 00c2f000 0000008 00c2f800 0000008 00c30000 0000008

Challenges of von Neumann Architecture

Moore's Law⁽¹⁾

The establishment Reliability coun degree of integration will be achieved with linear Present and future Interneted electronics is established In almost e ICs are widely applicable By integrated electronics, I mean Increasing the yield demonstrated h to integrated electronics in the linear are or nev technologies which are referred to There is no fundamental obstacle to achieving and level of production-low compared to that of distronics today as well as any additidevice yields of 100%. At present, packaging costs vable result in electronics functions suppli crete components-it offers reduced systems cost, so far exceed the cost of the semiconductor structure itself that there is no incentive to improve illo, fo ICs are increasingly p and in many systems improved performance has yields, but they can be raised as high as is ecohe reli been realized. ICs are more reliable nomically justified. No barrier exists comparable that to miniaturize electronics equipment to the thermodynamic equilibrium considerations ilure as the creasingly complex electronic functi Heat problem space with minimum weight. Sever Will it be possible to remove the heat generated taoish by tens of thousands of components in a single evolved, including microassembly individual components, thin-film silicon chipi 15 eat is a solvable issue OF THE COMPONENTS TED FUNCTION semiconductor integrated circuits. Moore' 13 12 Day of reckoning Impo Clearly, we will be able to build such component-10 Two-mil squares crammed equipment. Next, we ask under what 9 circumstances we should do it. The total cost of historic With the dimensional tolerances already being making a particular system function must be miniemployed in integrated circuits, isolated high-per **ICs are easy to manufacture** mized. To do so, we could amortize the engineerformance transistors can be built on centers two ing over several identical items, or evolve flexible thousandths of an inch apart. Such a two-mil square techniques for the engineering of large functions and they're getting smaller and can also contain several kilohms of resistance so that no disproportionate expense need be borne ICs are small by a particular array. Perhaps newly devised desmaller esigning ICs can be easy proponents onto integrated circuits', Electronics 38 (8). (1) Mo YEAR ZO

Linear circuitry

Integration will not change linear systems as radically as digital systems. Still, a considerable

Moore's Law⁽¹⁾

 The number of transistors we can build in a fixed area of silicon doubles every 12 ~ 24 months.



2015

CPU is important but...







Multicore processors







AMD Athlon 64 X2 (2005)2 cores



Nvidia Tegra 3 (2011)**5** cores





Intel Nahalem (2010)4 cores



Heterogeneous Computer Architecture













Storage

Performance gap between Processor/Memory





Heterogeneous Computer Architecture



Why should I care about "Computer Architecture"

What do you care when you're writing a program?



Algorithms Data Structures Computer Architecture Programming Languages User Interfaces

Demo (1)



Demo (2) — merge sort v.s. bitonic sort

Merge Sort $O(nlog_2n)$

void BitonicSort() { int i,j,k; for (k=2; k<=N; k=2*k) {</pre> for (j=k>>1; j>0; j=j>>1) { for (i=0; i<N; i++) {</pre> int ij=i^j; if ((ij)>i) { } } }

```
Bitonic Sort
 O(nlog_2^2n)
```

```
if ((i&k)==0 && a[i] > a[ij])
    exchange(i,ij);
if ((i&k)!=0 && a[i] < a[ij])
    exchange(i,ij);
```



CA2018 Iring Lecture



....









UC RIVERSITY OF CALIFORNIA





Thinking about the washlet







Or a Tesla



What's going to be in the class?

Heterogeneous Computer Architecture

Performance
Performance
measurement
What affects
performance
Amdahl's Law

• Metrics

Memory

- Memory
 - hierarchy
- Hardware
- optimizations
- Software
 optimizations

00023800

Processor

- Pipelining
- OoO Execution
- Branch
 - predictions
- Software
 - optimizations



Parallelism

Tentative Schedule

	Торіс	Reading
09/30/2019	Introduction	
10/02/2019	Performance Evaluation (I)	Chapter 1 Cramming More Components Onto Integrated Circuits, G.E. Moore, Proceedings of the IEEE 86(1):82-85, Jan 1998
10/07/2019	Performance Evaluation (II)	Twelve Ways to Fool the Masses When Giving Performance Results on Parallel Computers Amdahl's Law in the Multicore Era
10/09/2019	Memory Hierachy	Appendix B.1-B.4
10/14/2019		
10/16/2019		
10/21/2019	Memory Hierachy (II)	Chapter 2.1-2.3
10/23/2019	Memory Hierachy (III)	
10/28/2019	Basic Processor Design	Appendix C.1, Appendix C.2, Chapter 3.1
10/28/2019	Midterm Review	
10/30/2019	Branch prediction	Chapter 3.3 An analysis of correlation and predictability: what makes two-level branch predictors work Retrospective: a study of branch prediction strategies
11/04/2019	Midterm	
11/06/2019	Branch prediction	
11/13/2019	000 Scheduling	Chapter 3.4
11/18/2019	OOO Scheduling	K. C. Yeager, "The Mips R10000 superscalar microprocessor," in IEEE Micro, vol. 16, no. 2, pp. 28-41, April 1996. R. E. Kessler, "The Alpha 21264 microprocessor," in IEEE Micro, vol. 19, no. 2, pp. 24-36, March-April 1999.
11/18/2019	000 Scheduling	
11/20/2019	SMT	Chapter 3.11 Exploiting choice: instruction fetch and issue on an implementable simultaneous multithreading processor Using a user-level memory thread for correlation prefetching
11/25/2019	СМР	The case for a single-chip multiprocessor
11/27/2019	Dark Silicon	Dark silicon and the end of multicore scaling Single-ISA Heterogeneous Multi-Core Architectures: The Potential for Processor Power Reduction
12/02/2019	TPU, FPGA	In-Datacenter Performance Analysis of a Tensor Processing Unit A Configurable Cloud-Scale DNN Processor for Real-Time AI=
12/04/2019	Fin Sevi Wiect to	
12/09/2019	Final Exam	
	cnange	You need to complete the reading of H&P and papers

 Due	Note	
Homework #1	Move to 10/28/2019 as midterm review	
	Move to 11/18/2019	
Homework #2		
	7pm @ WCH 143	
 Homework #3		
	7pm @ WCH 143	
Homework #4		
Project		
 Check due dates here		

Learning eXperience



Most lectures today ...







VIe



Peer instruction

- I'll bring in activities to ENGAGE you in exploring your understanding of the material
 - Let you practice
 - Bring out misconceptions
 - Let us LEARN from each other about difficult parts.
- You will be GET CREDIT for your efforts to learn in class
 - By answering questions with a clicker (Poll Everywhere)
 - Answer 80% of the clicker questions in class, get 10% of your final grade •

Peer instruction

- Before the lecture You need to complete the required reading
- During the lecture I'll bring in activities to ENGAGE you in exploring your understanding of the material
 - Popup questions
 - Individual thinking use your clicker to express your opinion
 - Group discussion use your clicker to express your group's opinion
 - Whole-classroom discussion we would like to hear from you



ed <mark>reading</mark> you in exploring

nion up's opinion om you



Before lectures: reading quizzes

- Reading assignments from
 - Computer Architecture: A Quantitative Approach (6th Edition) by John Hennessy and David Patterson
 - Papers
- Reading quizzes:
 - On iLearn
 - Due before the lecture, usually once a week. Check the schedule on our • webpage
 - You will have two chances. We take the **average**
 - No time limitation until the deadline
 - No make up reading quizzes we will drop probably one or two lowest at least



About the time of the Lecture — Setup polling everywhere



Preview





Login through the app using UCRNetID@ucr.edu



Assignments and term project

- Assignments
 - The best way to prepare for midterm and final
 - Publish on the website
 - Submit through iLearn
- Project
 - C/C++ programming
 - Individual project
 - It's going to be a "contest" the winner will have a prize



Why papers?

No alternative facts

- Papers are written by authors who create/invent these artifacts
 - First-hand information
 - Not being cooked by media/press...
- Papers are reviewed based on originality
- Papers are reviewed by experts without conflict of interests



Papers give you insights!

- Papers contain design principles that are missing in your textbook or online documents
- You can apply these design principles and the skills of analyzing these principles to anywhere (e.g. you will surprisingly find how the paper you read next week affects software engineering)
- You can learn those whys for those proposed work



Industry cares

@intel.com> 寄給 h1tseng 👻		2011/2/15 🖄 🔸 🔻			
Hi Hung-Wei,					
I am very interested in your topic you	upresented yesterday. If possible, may liget a	copy of your presentation?			
Best Regards,	freeso 寄給 h1tseng I	ale.com 透過 cs.ucsd.edu			
	Hung-Wei				
	I just finished reading your paper "Understanding the Impact of Memory", very interesting information, do you have a PowerPothis paper?				

sa 寄給 h1tseng ▼	sap.com <u>透過</u> cs.ucsd.edu 2012/11/12 ☆ 🔨 🕤				
Hi Tseng,					
I have read your paper titled "Understanding the Impact of Power Loss on Flash Memory". It is an interesting piece of work. I would like to understand what specific tools did you use to observe the page-read and page-write behaviour at the FTL level. Did you use some sort of Flash simulator to get all the statistics about the numbers of page accesses and the energy consumption? My second question would be regarding FTL algorithms. Did you implement FTL for a real SSD or you used some kind of simulator and simulated the FTL algorithm?					
Thanks.	@huawei.com?	>			
	奇給 h1tseng -				
SAP Research	HI, Hung-wei,				
	from Huawei, and I a processing in ssds may be a promisir slides presented in the conference? I	im impressed by your ISCA 2016 present ig solution for future data centers. Would really appreciate your kindness. Thank y			
	Best regards,				





Make yourself more valuable

- Every top 20 CS MS/PhD program has their students reading papers in graduate-level classes
- You have to compete with them when you're on the market
- You need some context to prove that you're also geeky enough to be one of their colleagues

https://www.whitehouse.gov/the-press-office/2017/04/18/presidential-executive-order-buy-american-and

supersede or revise previous rules and guidance if appropriate, to p interests of United States workers in the administration of our immisystem, including through the prevention of fraud or abuse.

(b) In order to promote the proper functioning of the H-1B visa prog Secretary of State, the Attorney General, the Secretary of Labor, and Secretary of Homeland Security shall, as soon as practicable, sugge to help ensure that H-1B visas are awarded to the most-skilled or high petition beneficiaries.

Sec. 6. General Provisions. (a) Nothing in this order shall be constru impair or otherwise affect:



-hire-american	
rotect the	
gration	
	c
gram, the	e
l the	(
st reforms	`
ghest-paid	(
	(
ued to	

How to read research papers

• For each paper, you should identify the followings:

•	Why?	The most in
	 Why should we care about this paper? 	
	 What's the problem that this paper is trying to addres 	s?
•	What?	The second I
	 What has been proposed? 	you
	 Contributions of the paper 	
•	How?	The
	 How does the paper accomplish the proposed idea? 	want to im
	 How does the result perform? 	



nportant thing when you're reading/writing a paper

most important thing when 're reading/writing a paper

ey are important only if you plement the proposed idea

Recap & Brainstorm

- What are those related papers that you read before?
- Compare with those related papers and re-exam their **whys**, **whats** and • hows
- What will you propose if you're solving the same "why"?

Why is reading papers important

- As a researcher
 - You want to identify **important problems**
 - You want to know what has been accomplished don't reinvent the wheels
- As an engineer
 - You want to know if there is a solution of the design problems of your systems, applications
 - You want to know if you can apply the proposed mechanism
 - You want to know how to do it



Logistics

Course resource

- Lectures: MW 9:30a-10:50a, Boyce 1471
- Schedule, slides on course webpage: <u>https://www.escalab.org/classes/cs203-2019fa/</u>
- Discussion on piazza: <u>https://piazza.com/class/k12q0c5lo2v38k</u>
- Reading quizzes, homework submissions on iLearn: <u>https://ilearn.ucr.edu/</u>
- We do podcasting: TBA

Instructor — Hung-Wei Tseng

- Website: <u>https://intra.engr.ucr.edu/~htseng/</u>
- Office hour: MW 1:00p-2:00p @ WCH 406
- E-mail: htseng+CS203@ucr.edu
- BS/MS in Computer Science, National Taiwan University
- PhD in Computer Science, University of California, San Diego
- Research Interests
 - Intelligent storage devices
 - Non-volatile memory based systems
 - Near-data processing
 - Anything could accelerate applications



Teaching Assistant — Shixiong Qi

- Office hours: TuTh 9:30a-10:30a @ WCH 110
- E-mail: shixiong.qi@email.ucr.edu





Your tasks

- Login/discussion in iLearn and piazza. ٠
- Read the text before class!
 - **Computer Architecture: A Quantitative Approach (6th Edition)** by John Hennessy and David Patterson — previous editions are not supported
 - I'm not going to cover everything in class, but you are responsible for all the assigned text.
 - Papers
- Reading quizzes in iLearn (10%)
- Come to class (10%)
 - I will cover things not in the book.
 - 10% from clickers
- Homework throughout the course. (15%)
 - Help to practice the concepts from each topic
- Project (10%)
- Midterm (20%)
- Cumulative final (35%) •



• You can see your grades on iLearn.

UC San Diego	TritonEd	L Test Bur
		TritonEd Community
A 🔿 Teols		
CSE 141 - 15eng (S715)	Tools	
kome Page Information Quinzes Komework	Create and view Course Announcements	Groups Create and manage formal groups of students to callaborate on work.
Ciscussions Tools	Open Blackboard Help for Students	Jeumals Create and menogejournals that can be assigned to each user in a group jor the perposes of private
help Library Help Academic Integrity	Blogs Create and manage Mags for Caurses and Course Groups	My Grades
	.0000000	

- Errors in grading
 - If you feel there has been an error in how an assignment or test was graded, you have one week from when the assignment is return to bring it to our attention. You must submit (via email to the instructor and the appropriate TAs) a written description of the problem. Neither I nor the TAs will discuss regrades without receiving an email from you about it first.
- For arithmetic errors (adding up points etc.)
 - you do not need to submit anything in writing, but the one week limit still applies.

Academic Honesty

- Don't cheat.
 - Cheating on a test will get you an F in the class and no option to drop, and a visit with your college dean.
 - Cheating on homework means you don't have to turn them in any more, but you don't get points either. You will also take at least 25% penalty on the exam grades.
- Copying solutions of the internet or a solutions manual is cheating
 - They are incorrect sometimes
- Review the UCR student handbook
- When in doubt, ask.



2017 Fall @ NC State

LEAELE

You

2019 Summer II @ UCSD

2018 Spring @ NC State



2019 Summer I @ UCSD

2012 Summer @ UCSD

2017 Spring @ NC State



2016 Fall @ NC State

2019 Fall @ UCR





2014 Summer @ UCSD

Announcements

- Login piazza, iLearn
- Check our website
- Get "everywhere polling" ready
- Reading quiz due this Wednesday before class

Thank you!